

APPLICATION FOR UNITED STATES LETTERS PATENT  
FOR  
SYSTEM AND METHOD FOR CLOSED CAPTION DATA TRANSLATION

Inventors: Mary A. Flanagan  
Philip Jensen  
Douglas P. Chinnock

Assignee: Multilingual Media, Inc.  
Framingham, Massachusetts

Attorneys: Standley & Gilcrest LLP  
Attn: Carol G. Stovsky  
495 Metro Place South, Suite 210  
Dublin, Ohio 43017  
Telephone: (614) 792-5555  
Facsimile: (614) 792-5536

# SYSTEM AND METHOD FOR CLOSED CAPTION DATA TRANSLATION

Inventors: Mary A. Flanagan  
Philip Jensen  
Douglas P. Chinnock

## Technical Field

The present invention relates to transmission of closed caption data with broadcast signals. In particular, the present invention relates to translation of closed caption data from a source language to a target language.

## Background of the Invention

Despite the widespread access to television technology worldwide, language remains a barrier to broad dissemination of program content. More television content is developed in English than in any other language, yet English is spoken by only a tiny fraction of the world's population. Likewise, programming developed in other languages is inaccessible to speakers of English. A small amount of this content is translated by traditional means at high cost and with delays of weeks or even months. However, for television content that is perishable in nature, such as news, sports, or financial programs, there is no solution to broad distribution across languages. Such programming rapidly decreases in relevance over time, making the translation delays of weeks or more unacceptable. As a result, virtually all live television content goes untranslated, with different live programming developed specifically for each language market.

Live and time-sensitive television content is increasingly being delivered over the Internet in the form of streaming video. Broadband Internet access, a de facto

requirement for consumer access to streaming video, is being rapidly adopted by U.S. households. Market research suggests that by 2003, close to 9 million U.S. households will subscribe to a cable modem, up from 1.3 million at 1999 year-end. In Western Europe, exponential growth is predicted in the use of cable modems over the 1998-2003 time frame, and surveys are already showing that high speed access (ISDN or greater) is the predominant mode of Internet access. Regardless of the whether the delivery medium is a television set or an Internet-ready computer, language remains the critical barrier to widespread use of this broadcast content.

### Summary of the Invention

The present invention is a system and method for translating closed caption data. Closed caption data received from a television broadcast are translated, virtually in real-time, so that a viewer can read the closed caption data in his or her preferred language as the television program is broadcast. The present invention instantly localizes television program content by translating the closed caption data. The process of the present invention is fully automated, and may be used in conjunction with any machine translation system that has adequate performance to process translation in real-time to keep up with the program flow of caption data. A server supports real-time translation of eight television channels simultaneously, and translations are produced with less than a one-second delay. The server can produce either closed caption or subtitled output. An optional Separate Audio Program (SAP) may be added to the output that contains a computer generated speech rendering of one translation.

In accordance with the present invention, closed caption data is pre-edited to correct errors, recognize relevant text breaks, and enhance input quality to the machine translation system. For example, misspellings in the caption data are corrected before machine translation so that the machine translation system provides a correct translation from the source language to the target language. Incomplete sentences are detected and flagged or expanded so that the machine translation system provides a more accurate translation. The pre-editing process, which is unique to the present invention, results in high quality translations from commercially available machine translation systems. A unique text-flow management process further facilitates the processing and translating of text through the various components of the present invention.

#### **Brief Description of the Drawings**

Fig. 1 is a schematic diagram of the primary components for translation of streamed captions in accordance with an example embodiment of the present invention;

Fig. 2 is a schematic diagram of the primary components for translation of closed caption data with a combination decoder/subtitler device in accordance with an example embodiment of the present invention;

Fig. 3 is schematic diagram of the primary components for translation of time positioned captions in accordance with an example embodiment of the present invention;

Fig. 4 is a flowchart of the primary steps for closed caption text flow management in accordance with an example embodiment of the present invention; and

Fig. 5 is a flowchart of the primary steps for pre-editing of closed caption data in accordance with an example embodiment of the present invention.

#### Detailed Description of the Drawings

5 Referring to Fig. 1, a schematic diagram of the primary components for translation of streamed captions in accordance with an example embodiment of the present invention is shown. The program source 100 signal originates from a videotape recorder (VTR) or feed from a live cable or satellite signal. The program source 100 video, which may be in either National Television Systems Committee (NTSC) signal 104 format or National Association of Broadcasters (NAB) format consisting of video and closed caption (CC) data in the vertical blanking interval (VBI), is provided to both the CC decoder 106 and to the CC encoder 116 and another device 122. The other device 122 may be a subtitler that produces subtitles from translated text 114 received from the MT computer 110. Alternatively (or in addition), the other device 122 may be a text-to-speech (TTS) device (e.g., Lucent Technologies' "Lucent Speech Solutions" product) that synthesizes speech from the translated text 114. The synthesized speech from the TTS device 122 is placed into the Separate Audio Program (SAP) portion of the audio signal 102. Although Fig. 1 shows transmission of the NTSC signal 104 to the CC encoder 116 and the other device 122 (e.g., subtitler or TTS device), in alternative 20 embodiments of the present invention, the NTSC signal 104 may be transmitted to either the CC encoder 116 or the other device 122 and the MT computer may be adapted to send translated CC data 112 to a CC encoder 116 or translated text 114 to another device 122. Any type of signal that comprises closed caption data may be

directed to the MT computer 110 for translation. In addition to the NTSC signal, the present invention may also be used with the European NAB format program signal.

The CC decoder 106 extracts the CC codes (which consist of text, position, and font information) from the NTSC signal 104 and provides them to the MT computer 110 as a serial stream. In an example embodiment of the present invention, source language CC codes 108 may be transmitted from the CC decoder 106 to the MT computer 110.

The machine translation or MT computer 110 is a server that may be a Windows NT/2000 PC equipped with two serial ports. The MT computer 110 comprises machine translation (MT) software that performs automatic translation of human languages such as Transparent Language's Transcend SDK, Version 2.0. The MT software translates text from a first or source language to text in a second or target language. The MT software on the MT computer 110 translates the source language text stream or CC codes 108 from the CC decoder 106 to a target language. The target language may be any language (e.g., French, German, Japanese, or English) supported by the MT software on the MT computer 110. Then, the MT computer 110 merges the translated text stream with position and font information from the original CC codes. Resulting translated CC data 112 are transmitted to the CC encoder 116 as a serial stream. Resulting translated text 114 is transmitted to the other device 122 (e.g., subtitle or TTS device), also as a serial stream.

The CC encoder 116 combines the NTSC signal 104 or video portion of the program from the program source 100 and the translated CC data 112 from the MT computer 110 to produce a new, translated NTSC video signal 118. The translated

NTSC signal 118 is transmitted to the program destination 120. The final NTSC video signal 118, along with the audio signal 102 of the program source 100, is provided to the program destination 120, which may be a VTR or feed for a television or Internet broadcast.

5 Similarly, if the other device 122 is a subtitler, it combines the NTSC signal 104 or video portion of the program from the program source 100 and the translated text 114 from the MT computer 110 to produce a new, translated NTSC video signal 124. The translated NTSC signal 124 is transmitted to the program destination 126. The final NTSC video signal 124, along with the audio signal 102 of the program source 100, is provided to the program destination 126, which may be a VTR or feed for a television or Internet broadcast. In addition, or alternatively, if the other device 122 is a TTS device, it combines the audio signal 102 from the program source 100 to produce a SAP channel for the audio provided to the program destination 126.

Referring to Fig. 2, an example embodiment of the present invention is shown in which closed caption data is translated for a program destination in accordance with a combination decoder/subtitler device (e.g., an Ultech SG401). Audio signals 202 and NTSC signals 204 originate from a program source 200. The NTSC signal 204 or video signal (which consists of video and CC data) is transmitted from the program source 200 to an Ultech SG401 device that comprises a CC decoder 206 and subtitler 208.

20 The CC decoder 206 extracts the source language CC codes 210 which consist of text, position, and font information and provides them to the MT computer 212 as a serial stream. The MT computer 212, which comprises MT software as explained above, translates the source language CC codes 210 from the CC decoder 206. The MT

computer 212 merges the translated data with position and font information and provides the resulting translated text 214 to the subtitle 208, also as a serial stream. The subtitle 208 combines the video portion of the program from the program source and the translated text 214 from the MT computer 212. The result is a new translated NTSC signal 216 with translated subtitles. The final NTSC signal 216, along with the audio signal 202 from the program source 200, is provided to program destination 218 which may be a VTR or feed for a television or Internet broadcast. In addition, the translated text 214 may be processed by a text-to-speech (TTS) module (e.g., Lucent Technologies' "Lucent Speech Solutions" product) that synthesizes speech which is placed into the Separate Audio Program (SAP) portion of the audio signal provided to program destination 218.

Referring to Fig. 3, a schematic diagram of the primary components for translation of time positioned captions in accordance with an example embodiment of the present invention is shown. The program source 300 NTSC signals 304 are processed in two tape passes. The NTSC signals 304 originate from a VTR program source 300. The NTSC signals 304 from the VTR program source 300 consist of video and caption data in the VBI. The NTSC signals 304 are transmitted from the program source 300 to the CC decoder 306. In addition, timing codes 310 are sent from the VTR program source 300 to a MT computer 312. The MT computer 312 may be adapted to send translated CC data 314 to a CC encoder 318 or translated text 316 to another device 324 such as a subtitle or TTS device.

The CC decoder 306 extracts the source language CC codes 308 which consist of text, position, and font information and provides them to the MT computer 312 as a



serial stream. The MT computer 312 records, to a first file, the timing codes 310 and CC codes 308 for the entire program. The MT computer 312 then processes the first file to produce a second file with timing, translated data, position, and font information.

Next, a second pass of the program source tape 300 is made. On the second pass, the timing codes 310 are used by the MT computer 312 to determine when to send translated CC data 314 to the CC encoder 318 or the translated text 316 to the other device (e.g., subtitle or TTS device). The CC encoder 318 combines the video portion or NTSC signals 304 from the program source 300 and the translated CC data 314 from the MT computer 312. The result is a new translated NTSC signal 320 that is transmitted from the CC encoder 318 to a program destination 322.

Alternatively, or in addition, the other device 324 (e.g., subtitle or TTS device) combines the video portion or NTSC signals 304 from the program source 300 and the translated text 316 from the MT computer 312. The result is a new translated NTSC signal 326 that is transmitted from the other device 324 to a program destination 328.

In accordance with the present invention, the server, shown as the MT computer in Figs. 1, 2, and 3, in addition to MT software, may further comprise text flow management software and pre-editing software. Referring to Fig. 4, the primary steps for closed caption text flow management in accordance with an example embodiment of the present invention are shown. In an example embodiment of the present invention, the text flow management software, which is unique to the present invention, executes on a computer that also performs the machine translation. In an alternative embodiment of the present invention, the text flow management software and machine translation may execute on different computers that are connected or on a network. In

the first step 400, the text flow management software receives signals from a program source such as a television broadcast or videotape recorder. In the next step 402, an incoming stream of plain text that is present in the program source as text occurring in fields CC1, CC2, CC3, or CC4 in line 21 of the VBI is decoded or extracted using a closed caption (CC) decoder that passes the CC text to the text flow management software. An example device is the Ultech SG401 that operates as a closed caption decoder or subtitle character generator.

In the next step 404, the CC text is pre-edited to correct errors in closed captions, recognize relevant text breaks, and enhance input quality. The pre-edited text is translated from a source language to a target language using machine translation software in step 406. An example of machine translation software that may be used with the present invention is Transparent Language's Transcend SDK MT program.

In step 408, the target language text produced by the MT software is inserted into the video signal. It may be inserted as subtitles using the Ultech SG401 character generator or as closed captions replacing the original CC field or any of the fields CC1, CC2, CC3, or CC4 using CC encoder equipment from many suppliers. Finally, in step 410, the target language text is sent as a standard NTSC signal to a program destination for broadcast or recording to videotape recorder. The output of the text flow management process is a television program with translated closed captions or subtitles, depending on user preference. The closed captions or subtitles are properly synchronized with the program, either through producing the translations in real-time, or in some cases, through buffering the audio and video during the translation process, and reuniting audio, video, and text once the translations are complete.

Referring to Fig. 5, the primary steps for pre-editing of closed caption data in accordance with an example embodiment of the present invention are shown. The pre-edit software, which is unique to the present invention, solves several problems associated with real-time closed caption translation.

5 One problem with real-time closed caption translation is producing adequate quality translations, and doing so quickly enough so that the captions or subtitles keep pace with the live running video. Producing high quality translation of this unique text type involves several related problems. Captions that are produced on the fly for live programming such as news tend to have numerous misspellings and phonetic  
10 renderings of correct spellings. The misspellings result from the on-the-spot nature of the captioning task. Captioners who create the source language closed caption data must keep up with the real-time flow of speech. They are trained to use techniques such as phonetic spelling to quickly render proper names and other terms whose  
15 spelling cannot be determined instantly. The phonetic spellings often differ from common misspellings that occur when words are typed. Commercially available spell checking programs are not adequate for correcting these types of spellings. Because translation technology fails to recognize misspelled terms, the quality of the resulting translation is reduced. The present invention enhances the quality of the end result by pre-editing the closed caption data to recognize and correct this class of errors.

20 Another linguistic problem with real-time closed caption data is that a varying percentage of the text stream is complete sentences. This percentage often ranges from more than 85% in pre-written news broadcasts to as little as 20% in the unrehearsed speech of some speakers. The pre-editing techniques of the present

invention identify incomplete sentences before they are passed to the translation software. In some cases, incomplete sentences are expanded to structures that are easier for the translation software to handle. In other cases, they may simply be flagged so that they are not treated as full sentences by the translation software. In either case, the result is a more accurate translation of the closed caption data.

The vocabulary set for real-time broadcasts such as news presents yet another problem. In general, the vocabulary is broad and varied and therefore, requires ongoing additions to the machine translation software's dictionaries. The present invention addresses this problem by building specialized dictionaries according to topics. These specialized dictionaries are used in the translation process to produce higher quality translations. In addition to building dictionaries, topic changes are automatically identified during a program to determine which dictionary is appropriate for the context of the program. The building and automatic selecting of specialized dictionaries results in higher quality translations of closed caption data.

Referring to Fig. 5, the automated pre-editing process of the present invention comprises the following steps. First, in step 500, specialized dictionaries are developed according to topic. The context of a particular program may be very important in developing correct translations. The use of topic-based dictionaries for use by the machine translation software allows for more accurate translations. In the next step 502, the current program topic is identified to determine which dictionary should be used by the machine translation software. The topic may be identified by examining the frequency of the occurrence of certain key words or phrases. Other techniques may be

used to identify the appropriate topic. Once a dictionary is selected for the machine translation software, the process of translating incoming CC data may begin.

In step 504, phonetically based and other spelling errors occurring in the incoming text stream are corrected. Dictionaries that comprise phonetic spellings and associated correct spellings may be used to complete the correction of spelling errors. In the next step 506, sentence boundaries are identified and demarcated. In step 508, clause boundaries are identified and demarcated. After the sentence and clause boundaries are identified and demarcated, punctuation is added to the sentences and clauses, as appropriate in step 510. In step 512, ellipses appearing in the text stream are identified and text is inserted to complete the sentence. For unaccented text, accents are inserted where appropriate in step 514. In step 516, the speaker is identified based on CC position or voice print so the proper identifying information may be added to the output. Finally, in step 518, the pre-editing process checks for the end of the text stream to determine whether there is additional CC text to translate. If there is additional CC text to translate, the pre-editing process continues. Steps 502 to 516 are repeated for the incoming CC text.

The present invention translates closed caption data received from a live or taped television broadcast virtually in real-time so that a viewer can read the closed caption data in his or her preferred language during the broadcast. The present invention instantly localizes television program content by translating the closed caption data from a source language to a target language. The process of the present invention is fully automated, and includes a text flow management process and a pre-editing process that may be used in conjunction with any machine translation system. Various

modifications and combinations can be made to the disclosed embodiments without departing from the spirit and scope of the invention. All such modifications, combinations, and equivalents are intended to be covered and claimed.